MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2016/2017

DPS5018 – INTRODUCTION TO PROBABILITY AND STATISTICS

(Diploma in Information Technology)

18 OCTOBER 2016
9.00 a.m – 11.00 a.m
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **6** pages with **5** questions. Key formulae are given in the Appendix.
2. Answer **ALL** questions.
3. Write your answers in the answer booklet provided.
4. All necessary working steps must be shown.

Question 1

a)

> A survey of only 150 secondary students in SMK Danau finds that the student's average height is 5 feet 9 inches. From the survey, they make a conclusion that the height of student in SMK Danau is 5 feet 9 inches.

From the above statement, identify the sentence that best describes each type of statistics:

i) Descriptive (1 mark)

ii) Inferential (1 mark)

b) Identify the type of sampling used in each statement below: (4 marks)

No	Statement	Type of Sampling
Eg:	**Choosing volunteers for participation by drawing numbers from a jar.**	**Simple Random Sampling**
i)	If a researcher selected five schools at random in a city and then interviewed each of the teachers in those five schools.	
ii)	Selecting every third teenager leaving a movie theater to answer a survey about entertainment.	
iii)	Selecting students by the last digit of their student ID number to participate in a survey about their cafeteria.	
iv)	Listing the individuals by ethnic group and choosing a proportion from within each ethnic group at random.	

c) Define **Qualitative** and **Quantitative** variable. Provide one example for each variable. (4 marks)

[TOTAL 10 MARKS]

Continued...

Question 2

a) The following data shows the top 20 speed distribution of a sport at the 2010 Winter Olympics.

206	210	207	211	208
212	209	210	196	187
195	188	191	199	204
219	203	181	192	214

i) Construct stem-and-leaf diagram for the above data. (3 marks)

ii) From the diagram you construct in (i), find the median, mode and range. (5 marks)

iii) What is the percentage of player who performed faster than 210 seconds? (1 mark)

iv) Identify the shape of the speed distribution. (1 mark)

b) Students were asked to record the distance between their home and school. The distances were recorded in below table:

Student	Distance (km)
Ziana	0.4
Jackson	0.5
Benjamin	1.0
Nora	1.0
Joe	2.4
Nadia	2.7
Bella	3.0
Anthony	3.2
Amanda	4.1
Zain	4.2
Aida	4.4
Kenny	5.0
Johan	7.5

i) Calculate Q1, Q2 and Q3. (6 marks)

ii) Calculate the interquartile range, lower and upper limits. (3 marks)

Continued...

iii) A data point can be considered an "outlier" if it is above upper limit or below lower limit. Johan currently lives furthest from school. Would you consider his distance from school (7.5 km) to be "unusual"? Explain your answer. (1 mark)

[TOTAL 20 MARKS]

Question 3

a) At a local fast food restaurant, there are two types of burger that are served: chicken and fish. There are also two choices for drinks: milk or juice. Thirty percent of customers choose fish-burger while 70 percent choose chicken-burger. Forty percent of customers choose milk while 60 percent choose juice. Assume the drink choice is independent of burger choice.

i) Draw a tree diagram for this situation. (3 marks)

ii) What is the probability of a customer ordering Chicken-burger and Milk? (1 mark)

iii) What is the probability of a customer ordering Chicken-burger if he/she ordered Milk? (3 marks)

iv) What is the probability of a customer ordering Chicken-burger or Milk? (2 marks)

Use the following item prices to answer the **questions (v) to (vi)**:

Menu	Item	Cost RM)
Burger	Chicken	3
	Fish	4
Drink	Milk	1.50
	Juice	1

How much can the restaurant earn from a customer who orders:

v) a burger? (2 marks)

vi) a drink? (2 marks)

Continued...

b) A common blood test indicates the presence of a disease 95% of the time when the disease is actually present in an individual. Joe's doctor draws some of Joe's blood, and performs the test on his drawn blood. The results indicate that the disease is present in Joe.

Here is the information that Joe's doctor knows about the disease and the diagnostic blood test:

- One-percent of people have the disease (D), then $P(D) = 0.01$
- If the person is healthy (H), then $\mathrm{P}(H) = 1 - P(D) = 0.99$
- If a person has the disease (D), and the test is positive (T^+) then, $P(T^+|D) = 0.95$

i) Find $P(D')$. (2 marks)

ii) Find $P(D \cap T^+)$. (3 marks)

iii) Given that the blood test is positive for the disease is 5.9%, what is the probability that Joe actually has the disease? (2 marks)

[TOTAL 20 MARKS]

Question 4

a) In a Faculty of Arts, the following probability distribution was obtained for the number of students per semester taking Introductory Statistics course:

Number of students (x)	10	12	14	16	18
P ($X = x$)	0.1	0.15	0.3	0.25	0.2

i) Find the mean of this distribution. (2 marks)

ii) Find the probability if the number of student is less than 16. (3 marks)

b) In a group of children, the mean and variance of the number of children who own a LEGO toy are 7 and 3.5 respectively.

i) Find the value of probability, p and the number of sample, n. (3 marks)

ii) Find the probability if the number of children who own a LEGO toy is between 3 and 8 inclusively. (3 marks)

Continued...

c) The number of robbery reported every day at a police station have a Poisson distribution with mean of 2.

i) Find the standard deviation. (1 mark)

ii) Calculate the probability that number of robbery reported is more than 3. (3 marks)

d) The mass of apples plucked from an orchard follows a normal distribution with a mean of 640g and a standard deviation of 6g.

i) Find the mass that corresponds to the Z-score of 0.5. (2 marks)

ii) Calculate the probability that an apple chosen at random will have a mass less than 650g. (3 marks)

[TOTAL 20 MARKS]

Question 5

a) Data of shoe size, x and height, y for 10 students are presented in the table below.

Shoe size, x	Height, y
6	151
12	178
9	165
7	155
11	170
9	164
10	170
7	156
13	182
8	160

i) Find the regression line. **(Give your answer in 4 decimal places)** (12 marks)

ii) Obtain the linear correlation coefficient, r, and interpret the relationship. (3 marks)

Continued...

b) A survey had been done on starting salary between 35 candidates of fresh graduate and 35 experienced employees. It has been claimed that the mean salary of fresh graduate is always lower than the mean salary of experienced employee. The statistics are summarized as below.

Fresh graduate	Experience employee
$\bar{x}_1$ = RM2400	$\bar{x}_2$ = RM4560
σ_1 = RM7600	σ_2 = RM14200

i) State the null and alternative hypothesis. (2 marks)

ii) Obtain the *Z*-score. (2 marks)

iii) Should the claim be accepted at 5% significance level? (3 marks)

c) The Cubic Sdn Bhd claimed that 45% of its customers are satisfied with its customer service. However, a new executive wanted to test this claim and selected a random sample of 100 customers. It was found that only 38% were satisfied. At 1% significance level, does the sample give enough evidence to accept the claim? (8 marks)

[TOTAL 30 MARKS]

End of Page.

APPENDIX – KEY FORMULAE

- **Mean for ungrouped data**

$$\bar{x} = \frac{\sum x_i}{n} \qquad or \qquad \mu = \frac{\sum x_i}{N}$$

- **Mean for grouped data**

$$\bar{x} = \frac{\sum x_i f_i}{n} \qquad or \qquad \mu = \frac{\sum x_i f_i}{N}$$

Where x is the midpoint and f is the frequency of the class

- **Variance for ungrouped data**

$$s^2 = \frac{\sum {x_i}^2 - \frac{(\sum x_i)^2}{n}}{n-1} \qquad or \qquad \sigma^2 = \frac{\sum {x_i}^2 - \frac{(\sum x_i)^2}{N}}{N}$$

- **Variance for grouped data**

$$s^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n-1} \qquad or \qquad \sigma^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{N}}{N}$$

Where x is the midpoint and f is the frequency of the class

DISCRETE RANDOM VARIABLE, BINOMIAL DISTRIBUTION, AND POISSON DISTRIBUTION

- Mean of a discrete random variable x : $\mu = \sum xP(X = x)$
- Standard deviation of a discrete random variable x : $\sigma = \sqrt{\sum x^2 P(X = x) - \mu^2}$
- Binomial probability formula : $P(X = x) = \binom{n}{x} p^x q^{n-x}$
- Poisson probability formula : $P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$

- Key formula for Binomial Probability Distribution :
  - If $p \leq 0.5$
    i. $P(X = r) = B(r) - B(r-1)$
    ii. $P(X \geq r) = 1 - B(r-1)$
    iii. $P(X \leq r) = B(r)$
    iv. $P(X > r) = 1 - B(r)$
    v. $P(X < r) = B(r-1)$
    vi. $P(a \leq X \leq b) = B(b) - B(a-1)$

  - If $p > 0.5$
    i. $P(X = r) = B(n-r) - B(n-r-1)$
    ii. $P(X \geq r) = B(n-r)$
    iii. $P(X \leq r) = 1 - B(n-r-1)$
    iv. $P(a \leq X \leq b) = B(n-a) - B(n-b-1)$

- Key formula for Poisson Probability Distribution :
  - If $\mu = \lambda$
    i. $P(X = r) = Poi(r) - Poi(r-1)$
    ii. $P(X \geq r) = 1 - Poi(r-1)$
    iii. $P(X \leq r) = Poi(r)$
    iv. $P(X > r) = 1 - Poi(r)$
    v. $P(X < r) = Poi(r-1)$
    vi. $P(a \leq X \leq b) = Poi(b) - Poi(a-1)$

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $z = \frac{x-\mu}{\sigma}$

REGRESSION

- For the **least square regression**, $\hat{y} = a + bx$

where, $SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$, $SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$

and, $b = \frac{SS_{xy}}{SS_{xx}}$, $a = \bar{y} - b\bar{x}$

HYPOTHESIS TESTING

- **ONE population**

Mean:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Proportion:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

- **TWO population**

Mean:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

Proportion:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$